

**JP6259198**

Biblio

Page 1

Drawing

**DISK DEVICE SYSTEM**

Patent Number: JP6259198  
Publication date: 1994-09-16  
Inventor(s): MATSUMOTO JUN; others: 03  
Applicant(s): HITACHI LTD  
Requested Patent: ☐ JP6259198  
Application Number: JP19930048051 19930309  
Priority Number(s):  
IPC Classification: G06F3/06; G11B19/02; G11B20/10;  
G11B20/12  
EC Classification:  
Equivalents: JP3250861B2

**Abstract**

**PURPOSE:** To attain the high speed of a command processing in disk device system.  
**CONSTITUTION:** A command scheduler 32 which receives a command from a host computer is equipped with a command waiting queue 33 in which arriving commands wait, command rearrangement queue 34 used for the rearrangement of the commands, queue control part 35 which controls the queue, queue size control means 36 which controls the size of the queue, and execution sequence optimization control part 37. The execution sequence optimization control part 37 is equipped with a continuous area access command rearranging means 38 which rearranges the commands for performing an access to continuous areas, and a multiple execution possibility rearranging means 39 which rearranges the commands to be multiple-executed. The commands transmitted from a host computer are rearranged so that the access to the continuous areas can be performed by the address of data, and so that the multiple execution can be attained.

Data supplied from the esp@cenet database - l2

**BEST AVAILABLE COPY**



1

## 【特許請求の範囲】

【請求項1】 ディスク装置と上記ディスク装置を制御するディスク制御装置とを有し、外部からの、データの読み出しまたは書き込みのコマンドを受けて、上記データの読み出しまたは書き込みを行うディスク装置システムにおいて、

上記ディスク制御装置は、

上記コマンドを受信する受信手段と、

上記コマンドを蓄積する第一コマンドキューと、

上記蓄積されたコマンドの実行時間の総和を短縮するように上記コマンドの実行順序を変更する実行順制御手段とを有することを特徴とするディスク装置システム。

【請求項2】 請求項1記載のディスク装置システムにおいて、

上記実行順制御手段は、上記コマンドが処理対象とするデータのアドレスを調べ、アドレスが連続するコマンドを抽出し、上記アドレスが連続するように、上記コマンドの実行順序を変更することを特徴とするディスク装置システム。

【請求項3】 請求項1記載のディスク装置システムにおいて、

上記ディスク装置システムは、複数のディスク装置を有し、

上記実行順制御手段は、上記コマンドが処理対象とするディスク装置を調べ、異なるディスク装置を処理対象とするコマンドが連続するように、上記コマンドの実行順序を変更し、

異なるディスク装置が並行して処理をすることを特徴とするディスク装置システム。

【請求項4】 請求項1記載のディスク装置システムにおいて、

上記実行順制御手段は、

上記コマンドが処理対象とするデータのアドレスを調べ、アドレスが連続するコマンドを抽出し、上記アドレスが連続するように、上記コマンドの実行順序を変更し、

次に、上記連続するアドレスを有する複数のコマンドは1つのコマンドとみなして、上記コマンドが処理対象とするディスク装置を調べ、異なるディスク装置を処理対象とするコマンドが連続するように、上記コマンドの実行順序を変更することを特徴とするディスク装置システム。

【請求項5】 請求項1、2、3または4記載のディスク装置システムにおいて、

予め定められた条件を満たしたときに、上記第一コマンドキューにコマンドを蓄積することを停止するキューサイズ制限手段と、

制限されて、上記第1コマンドキューに蓄積されなくなったコマンドを蓄積する第二コマンドキューと、

第一コマンドキューに蓄積したコマンドの並び変えが終

2

了後、上記コマンドを送出し、次に、第二コマンドキューにあるコマンドを第一コマンドキューに送るキュー制御手段とを有することを特徴とするディスク装置システム。

【請求項6】 請求項1記載のディスク装置システムにおいて、

上記キューサイズ制限手段は、第一コマンドキューにコマンドを蓄積することを、コマンド数が一定値に達した場合、または一定時間が経過した場合、または特定のコマンドを外部から受信した場合に停止することを特徴とするディスク装置システム。

## 【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、ハード・ディスク、光ディスク及び半導体ディスク等のディスク装置を使用するディスク装置システムにおける、大容量データの高速書き込み、読み出し処理に関する。複数ディスク装置で構成されるディスクアレイのコマンド処理にも適する。

【0002】

【従来の技術】 ファイルシステムの大容量化とトランザクション性能の向上が求められる中、大量のデータを高速に処理する一方式として特開平2-236714号公報に開示されるアレイ型ディスク駆動機構システム及び方法がある。この方式を図10を用いて説明する。

【0003】 図10において、10はRead/Writeコマンドを発行する上位計算機、11はデータの分配・集合制御を行うアレイ・コントローラである。11の中で、12は上位計算機とのコマンド及びデータの授受を行う上位計算機I/F部、13はタスク実行管理を行うタスク制御部、14はコマンド多重処理の最適化を行うコマンド処理管理部、及び上位計算機12からのコマンドをディスクアレイのコマンドに再編成するアドレス変換や各ディスク装置へのコマンド生成、発行、終了処理を行うコマンド処理部、15は割込み処理部、16～20はディスクとのコマンド及びデータ授受を行うディスク装置I/F部、21～25はデータの書き込み、読みだしを行うディスク装置である。なお、ここで扱うコマンドはSCSIコマンドであり、それ自体がコマンドの種類、論理ブロックアドレス（以下LBAと記す）、転送長の3つの要素を持っている。

【0004】 まず、図10に示すアレイ・コントローラ11によって複数のドライブ装置21～25を多重動作（並行に動作させること）させる。上位計算機10から転送されるコマンド・データは、アレイ・コントローラ11のコマンド処理管理部15にて、各ディスク装置に所定の単位で振り分けるためにアドレス変換される。各コマンドをファーストイン・ファーストアウト方式（以下FIFO方式と記す）により到着順にタスク制御管理部13や割込み処理部15の管理の下で、ディスク装置I/F部16～20を介して、各ディスク装置21～2

5に書き込む。コマンド・データの読み出しの場合には、各ディスク装置21～25から読み出された分割されたコマンド・データをアレイ・コントローラ11で1つのデータに統合し、上位計算機10に送出する。

【0005】

【発明が解決しようとする課題】かかる従来の方法においては以下の課題がある。

【0006】すなわち、従来のコマンドを到着順に処理するFIFO方式では、上位計算機が複数ディスクに対してRead/Writeコマンドを実行する場合、連続データをアクセスする複数のコマンド列中に、該連続データに無関係なデータをアクセスするコマンドが混在するような場合、その混入毎にヘッドシークが発生し、著しく連続データ転送特性を低下させるという課題がある。

【0007】また、あるコマンドがディスク装置0とディスク装置1の2台のディスクを使用するような場合、ディスク装置0がアクセス待ちをしていると、ディスク装置1は未使用にもかかわらず、そのコマンドに確保されたままになっているために、後から到着するコマンドがディスク装置1を利用することができず、多重動作が制限されるという課題がある。

【0008】本発明の目的は、コマンドの実行順序を変更して、処理の高速化を実現したディスク装置システムを提供することである。

【0009】

【課題を解決するための手段】上記の目的を達成するため、ディスク装置と上記ディスク装置を制御するディスク制御装置とを有し、外部からの、データの読み出しまたは書き込みのコマンドを受けて、上記データの読み出しまたは書き込みを行うディスク装置システムにおいて、上記ディスク制御装置は、上記コマンドを受信する受信手段と、上記コマンドを蓄積する第一コマンドキューと、上記蓄積されたコマンドの実行時間の総和を短縮するように上記コマンドの実行順序を変更する実行順制御手段とを有することとしたものである。

【0010】

【作用】コマンドが第一コマンドキューに入り、上記蓄積されたコマンドの実行時間の総和を短縮するように上記コマンドの実行順序を変更することにより、高速化が図れる。

【0011】たとえば、実行順制御手段が第一コマンドキューのコマンド群について、連続領域アクセス可能性を考慮して並び替え、並び替えが終わったコマンドから実行を開始することで、連続領域アクセスデータをまとめて処理することができる。

【0012】これにより、分断された連続領域アクセスデータをまとめて1つのコマンド群として処理することで、ディスクヘッドのシーク等を減らすことができ、高速化を実現することである。

【0013】また、第一コマンドキューのコマンド群について多重実行可能性を考慮して、並び替え、並び替えが終わったコマンドから実行を開始することで、コマンドを多重実行することができる。

【0014】これにより、複数のディスク装置が極力多重実行可能なように、コマンドを並び替えられるので、ディスクの稼働率を上げ、1コマンドあたりの平均処理時間を短縮することができる。

【0015】

【実施例】本発明の第一実施例を図1～4、11を用いて説明する。

【0016】図11に示す本ディスク装置システム（ディスクアレイ）は、Read/Writeコマンドを発行する上位計算機10から、コマンドを受けてデータの分配・集合制御を行うアレイ・コントローラ11（ディスク制御装置）と、データの書き込み、読みだしを行うディスク装置21～25とを有する。アレイ・コントローラ11は、上位計算機とのコマンド及びデータの授受を行う上位計算機I/F部12と、タスク実行管理を行うタスク制御部13と、コマンド多重処理の最適化を行うコマンド処理並びに管理、及び上位計算機10からのコマンドをディスクアレイのコマンドに再編成するアドレス変換や各ディスク装置へのコマンド生成、発行、終了処理を行うコマンド処理及び管理部141と、割り込み処理部15と、ディスクとのコマンド及びデータ授受を行うディスク装置I/F部16～20とを有する。なお、ここで扱うコマンドはSCSIコマンドであり、それ自体がコマンドの種類、論理ブロックアドレス（以下LBAと記す）、転送長の3つの要素を持っている。

【0017】まず、図11に示すアレイ・コントローラ11によって複数のドライブ装置21～25を多重動作（並行に動作させること）させる。上位計算機10から転送されるコマンド・データは、アレイ・コントローラ11のコマンド処理管理部15にて、各ディスク装置に所定の単位で振り分けるためにアドレス変換される。各コマンドをファーストイン・ファーストアウト方式（以下FIFO方式と記す）により到着順にタスク制御管理部13や割り込み処理部15の管理の下で、ディスク装置I/F部16～20を介して、各ディスク装置21～25に書き込む。コマンド・データの読み出しの場合には、各ディスク装置21～25から読み出された分割されたコマンド・データをアレイ・コントローラ11で1つのデータに統合し、上位計算機10に送出する。

【0018】図1は、コマンド処理及び管理部141の詳細を表わした図である。

【0019】コマンド処理及び管理部141は、上位計算機からのRead/Writeコマンドを受信する上位計算機コマンド受信部30（受信手段）と、Read/Writeコマンドの持つLBAと転送長から、上位計算機が送出したコマンドをアレイディスクのコマンド

に再編成するアドレス変換部31と、Read/Writeコマンドの実行順を並び変えるコマンド・スケジューラ32とを有する。

【0020】コマンド・スケジューラ32は、到着コマンドが待機するコマンド待機キュー33（第2コマンドキュー）と、コマンドの並び変えに用いるコマンド並び変えキュー34（第1コマンドキュー）と、コマンド待機キュー33、コマンド並び変えキュー34を制御するキュー制御部35（キュー制御手段）と、キューのサイズを制限するキューサイズ制限手段36と、実行順最適化制御部37（実行順制御手段）とを有する。

【0021】コマンドの実行順最適化制御部37は、連続領域をアクセスするコマンドを並び変える、連続領域アクセスコマンド並び変え手段38と、多重実行可能なコマンドを並び変える多重実行可能性並び変え手段39とを有する。

【0022】図2のフローチャートに基いて、図1の全体動作を説明する。

【0023】図1に示す上位計算機コマンド受信部30はRead/Writeコマンドを受信し（図2フローチャート200）、アドレス変換部31はアレイディスクにデータを分配するアドレス変換をコマンドに施す（同202）。コマンド・スケジューラ32はコマンドを取り入れ、それを到着順にコマンド待機キュー33にセットする（同204）。コマンドはコマンド待機キュー33を通過して、コマンド並び変えキュー34に入る。ある条件を設けることによって、キューサイズ制限手段36はコマンド待機キューのコマンド通過を停止する。したがって、コマンド並び変えキュー内のコマンド数は限定される（同206）。尚この条件には、例えばタイマーによってある一定時間経過後にキューを制限するタイミングを送出する方法及び、コマンド並び変えキューに到着するコマンド数をあらかじめ限定しておく方法がある。また、これから到着するコマンドは、コマンド並び変えキュー34で前に到着したコマンドの並び変えが終了するまで、コマンド待機キュー33で待機する。コマンド並び変えキュー内のコマンドについては到着順を前提に（同208）、初めに連続領域アクセス可能性（同210）、次に多重実行可能性（同212）の有無をチェックし、両方存在する場合は連続領域アクセス可能性を優先的に考慮してコマンドを並び変え、連続領域をアクセスしないコマンドについてののみ多重実行可能性を考慮してコマンドを並び変える。またどちらかが存在する場合は、独立してコマンドの並び変えを実施する（同214）。また、並び変えが終わったコマンドから実行を開始する（同216）。コマンド並び変えキューが空になったら（同218）、コマンド待機キュー33のコマンド通過停止を解除し、待機していた新しいコマンドをコマンド並び変えキュー34に入れ、以下これまでの動作が繰り返される。

【0024】第一実施例として多重実行可能性のみを考慮した場合を説明する。

【0025】図3は図1のコマンド・スケジューラ32に当たる。図3において、33は到着コマンドが待機するコマンド待機キュー、34はコマンドの並び変えに用いるコマンド並び変えキュー、40は33、34のキューを制御するキュー制御部、41はキューのサイズを制限するタイミングを決定するタイマー、42はアドレスのビット・パターン作成部である。43は先頭コマンド及び並び変えが終了したコマンドのビット・パターンを読み取る第一ビット・パターン読み取り部、44はまだ並び変えていないコマンドのビット・パターンを読み取る第二ビット・パターン読み取り部、45は43、44で読み取ったビット・パターンを比較する比較器である。

【0026】図4のフローチャートに基いて、図3全体の動作を説明する。

【0027】図3に示すコマンド・スケジューラはアドレス変換されたRead/Writeコマンドを取り入れ（図4のフローチャート300、302）、それを到着順にコマンド待機キュー33に入れる。この時、ビット・パターン作成部42はコマンド固有のLBAを元に当該コマンドが使用するディスク装置を示すビット・パターンを作成する。ここでビット・パターンは使用するディスク装置を1、使用しないディスクを0で表わし、図10のディスク装置の左からそれぞれ対応させたビット列である。例えば、0から2番のディスク装置を使用し、3、4番のディスク装置を使用しない場合、ビット・パターンは11100となる。また、ビット・パターン作成部42は作成したビット・パターンをコマンドに対応づけて保持する（同304）。そこで、第一ビット・パターン読み取り部43はキューの先頭コマンドのビット・パターンを読み取り、反転する（同308）。反転したビット・パターンに1がある場合、つまり空いているディスク装置がある場合は、第二ビット・パターン読み取り部44が、並び変える残りのコマンドのビット・パターンを読み取り、比較器45がそれを第一ビット・パターン読み取り部のビット・パターンとビット毎に比較し（同310）、一致する1の最も多いコマンドを選択する。その後キュー制御部40はそのコマンドを先頭コマンドの後に並び変える（同312）。また、並び変えたコマンドから実行が開始され（同314）、コマンド並び変えキューが空になったら（同316）、コマンド待機キュー33のコマンド通過停止を解除し、待機していた新しいコマンドをコマンド並び変えキュー34に入れ、以下これまでの動作が繰り返される。ビット・パターンの使用はディスクアクセスの多重度を上げるために、使用するディスク装置ができるだけ重ならないようにコマンドを選択する並び変え方式において、簡便で有効な手段である。

【0028】具体例として、図3に示すような10個のRead/Writeコマンドがコマンド並び替えキュー34に到着している場合のコマンド並び替え動作を示す。ここでC#はコマンドの到着番号を、(####)は各コマンドのビット・パターンを表わしている。まず、第一ビット・パターン読み取り部43はC1のビット・パターン11100(ディスク0、1、2番を使用する)を読み取り、この情報を反転して00011を得る。次に第二ビット・パターン読み取り部44がC2からC10までのビット・パターンを読み取り、比較器45がこの反転ビット・パターンと一致もしくは一致する1の多い(0は全て対応していなければならない)コマンドを検索する。この例では対応するコマンドがないので、C2が2番目のコマンドになる。こんどは第一ビット・パターン読み取り部43がC2のビット・パターン00101を読み取り、このビット・パターンを反転して11010を得る。ここで、C1の時と同様に、第二ビット・パターン読み取り部44がC3からC10までのビット・パターンを読み取り、比較器45が対応するコマンドを検索する。すると、11000のC9が見つかり、実行順を入れかえ、C2の次に配置する。このC2とC9は多重実行である。さらに多重実行できる可能性を判断するため、キュー制御部40がC2とC9のEOR(両方0のビットをマスクする)00010を作り、第一ビット・パターン読み取り部43がこれを読み取る。また、第二ビット・パターン読み取り部42がC3からC8、C10のビット・パターンを読み取り、比較器43が対応するコマンドを検索するが、ないのでキュー制御部35がC3を4番目のコマンドに並び替える。以下この操作が繰り返される。並び替えた結果を図9(a)に示す。

【0029】この具体例によれば、各コマンド処理を極力多重動作で行うため、たとえば上記例の場合、コマンド処理時間を全てのコマンドで等しく1と仮定すると、実行順を変える前と変えた後では、10コマンドの処理時間合計は本発明未使用時の10が、本発明使用により6に削減され、40%の改善効果がある。したがって、この多重実行最適化によって、コマンドのI/Oスループットを上げることができる。

【0030】また、キューを2つ設けることで、並び替えたコマンドが全て実行されてから新しいコマンドの並び替えが行われるため、キューを1つしか用いなかった従来の場合に見られたコマンドの沈み込み(例えば、優先度が低いコマンドは、いつまでもキューの中に実行されないで残る現象)を解消することができる。

【0031】第二実施例として連続領域アクセス可能性のみを考慮した場合を説明する。

【0032】図5のコマンドスケジューラ532は図1のコマンド・スケジューラ32に相当する。図5において、33は到着コマンドが待機するコマンド待機キュー

一、34はコマンドの並び替えに用いるコマンド並び替えキュー、50は33、34のキューを制御するキュー制御部、41はキューのサイズを制限するタイミングを決定するタイマー、51は連続領域判定アドレス作成手段である。537は、実行順最適化制御部である。51の中で52は先頭コマンド及びまだ並び替えが終了していないコマンドの内、コマンド・ナンバーの若いもののアドレスを読み取る第一アドレス読み取り部、53は52に入るコマンド転送長を読み取る転送長読み取り部、54は第一アドレス読み取り部52で読み取ったコマンドの先頭アドレスに転送長を加えて、連続する領域のアドレスを算出する加算機、56はまだ並び替えていないコマンドのアドレスを読み取る第二アドレス読み取り部、57は51と56のアドレスを比較する比較器である。

【0033】図6のフローチャートに基いて、図5全体の動作を説明する。

【0034】図5に示すコマンド・スケジューラはアドレス変換されたRead/Writeコマンドを取り入れ(図6のフローチャート400、402)、それを到着順にコマンド待機キュー33に入れる。コマンドは最初コマンド待機キューを通過してコマンド並び替えキュー34に入り、タイマー41が一定時間経過後、並び替えるキューサイズを制限する(同404)。この後到着する新しいコマンドはコマンド待機キューで待機する。次に第一アドレス読み取り部52はコマンド並び替えキュー先頭のコマンドが持つアドレスを読み取り、転送長読み取り部53はその転送長を読み取り(同406)、読み取ったアドレスに転送長を加算器54で加えることで、連続する領域のアドレスを算出する(同408)そして、第二アドレス読み取り部55は並び替える残りのコマンドが持つアドレスを読み取り、比較器56がそれを連続領域判定アドレス作成手段51と比較し(同410)、一致するコマンドを先頭コマンドの後に並び替える(同412)。また、実行順が決定したコマンドから実行が開始され(同414)、コマンド並び替えキューが空になったら(同416)、コマンド待機キュー33のコマンド通過停止を解除し、待機していた新しいコマンドをコマンド並び替えキューに入れ、以下これまでの動作が繰り返される。

【0035】具体例として、図5に示すような10個のRead/Writeコマンドがコマンド並び替えキュー34に到着している場合のコマンド並び替え動作を示す。C#はコマンドの到着番号を、(##)は左がコマンドの持つアドレス、右が転送長を表わしている。まず、連続領域判定アドレス作成手段51の中の第一アドレス読み取り部52がC1のアドレス0を読み取り、転送長読み取り部53がC1の転送長16ブロック(1ブロック=512B)を読み取り、加算機54はアドレス0に16を加えて16を算出する。次に、第二アドレス

読み取り部55がC2からC10までのアドレスを読み取り、比較器56が連続領域判定アドレス作成手段51で算出された値16と等しいアドレスを持つコマンドを検索する。すると、C3が見つかるので、キュー制御部50がC3をC1の後に並び変える。さらに連続領域アクセスできる可能性を判断するため、第一アドレス読み取り部52がC3のアドレス16を読み取り、転送長読み取り部53がC3の転送長16ブロックを読み取る。前述と同様な操作でコマンドを検索するとC6が見つかる。ここまでは、コマンドはC1、C3、C6の順に並び変えられる。この後C6について再び連続領域アクセスの可能性を判断すると、今度は対応するコマンドがないため、C2がC6の後に並び変えられる。以下この操作が繰り返される。並び変えた結果を図9(b)に示す。

【0036】この具体例によれば、1つのコマンドに対するアクセス回数を1とすると、実行順を変える前では連続領域アクセスが可能なコマンド群はC8、C9のみで、アクセス回数は9であるが、実行順を変えた後では連続領域アクセスが可能なコマンド群として、C1、C3、C6及びC5、C7、C10も加わるので、アクセス回数は5である。一般的に、十分たくさんのコマンドを連続化することでディスク列全てにアクセスし、ヘッドシーク時間を短縮することができる。また長いデータに他のコマンドが混入する場合のヘッドシーク時間増加を防止する効果もある。

【0037】さらに、長いデータをアクセスする場合、ディスク装置の持つプリフェッチ・バッファをきかせることができる。すなわち、アドレスが連続しない短いデータを有するコマンドの場合、コマンドごとにプリフェッチ・バッファの内容を書き換える必要が生じる。書き換えの回数が増えることにより処理時間が延びる。

【0038】本実施例はディスクアレイのみならず、単体のディスク装置に対しても有効なものである。

【0039】第三実施例として多重実行可能性のみを考慮した場合を説明する。

【0040】図7は図1のコマンド・スケジューラ32に当たる。図7において、61~70はk台のディスク装置の組合せに対応した各コマンド到着キュー(n台のディスク装置からk台のディスク装置を選ぶ全ての組合せの数だけ用意したキュー)、71~80はコマンド到着キューがコマンドを実行キューに送出したことを示すフラグ、81は並び変えを終了したコマンドが実行を待つコマンド実行キューである。60はコマンド到着キュー61~70及びフラグ71~80の制御をするキュー制御部、42はアドレスのビット・パターン作成部、82はコマンドの並び変えテーブル参照手段である。

【0041】次に図8のフローチャートに基いて図7全体の動作を説明する。

【0042】図7に示すコマンド・スケジューラ732はアドレス変換されたRead/Writeコマンドを取り入れ(図8のフローチャート500、502)、ビット・パターン作成部42はコマンド固有のLBAを元に当該コマンドが使用するディスクを示すビット・パターンを作成する。次に、キュー制御部60でそのビット・パターンに対応した各コマンド到着キュー61~70に振り分けられる(同504、506)。コマンド到着キュー先頭のコマンドが全てそろったら、キュー制御部60がコマンド・ナンバーの最も若いコマンドを選択し(同508)、そのコマンドをコマンド実行キュー81にセットし、フラグ71~80のうちコマンドを送出したキューのフラグを立てる(同510)。コマンドを送出したキューと対応するキューをコマンド並び変えテーブル参照手段82で参照し(同512)、対応するコマンドがなければ次に若いコマンド・ナンバーのコマンドをコマンド実行キュー81に送出し、フラグ71~80のうちコマンドを送出したキューのフラグを立てる(同510)。対応するコマンドがあれば、そのコマンドをコマンド実行キュー81に送出し、フラグ71~80のうちコマンドを送出したキューのフラグを立てる(同514)。以上の操作をコマンド到着キュー61~70の全てのフラグ70~80が立つまで行う(同516)。コマンド実行キュー81に並び変えたコマンドが到着しだい、コマンドを実行し(同518)、コマンド実行キュー81が空になったら(同520)、各コマンド到着キュー61~70の全てのフラグ71~80を解除する(同522)。そして、再びコマンド到着キュー61~70の中で、最も若いコマンド・ナンバーのコマンドがコマンド実行キュー81に送出され、以下これまでの動作が繰り返される。

【0043】具体例として、図7に示すような2台のディスク装置にアクセスする複数のRead/Writeコマンドがコマンド到着キュー61~70に到着している場合のコマンド並び変え動作を示す。ここで、C#はコマンドの到着番号を表わしている。732は、コマンドスケジューラ、61~70は到着コマンドが待機するコマンド待機キュー、60は61~70のキューを制御するキュー制御部、737は、実行順最適化制御部である。まず、コマンド到着キュー61~70の先頭にコマンドが全てそろった状態で、キュー制御部60がコマンド・ナンバーの最も若いC1をコマンド実行キュー81にセットし、フラグ71を立てる。C1は0番と1番のディスク装置を使用するので、このコマンドと多重実行可能なコマンドをコマンド並び変えテーブル参照手段82において対応づける(図8のテーブル参照)。すると、3番と4番のディスク装置を使用するコマンドが適切ということから、キュー制御部60がC7をコマンド実行キューに送出し、フラグ80を立てる。次に、キュー制御部60がフラグのまだ立っていないコマンド到着

キューの内、コマンド・ナンバーの最も若いC2をコマンド実行キューに送出し、フラグ76を立てる。C7の操作と同様にC2と多重実行可能なコマンドを検索するとC10が見つかり、キュー制御部60がC10をコマンド実行キュー81に送出し、フラグ79を立てる。以下この操作が繰り返される。並び変えた結果を図9(c)に示す。

【0044】この具体例によれば、全てのコマンド処理を多重動作で行うため、コマンド処理時間を全て等しく1と仮定すると、実行順を変える前では10コマンドの処理時間が、実行順を変えた後では5となる。一般に実行順を変える前は、コマンド処理時間はコマンドの並び方によって5～10の範囲（完全に多重実行する場合から1コマンドも多重実行しない場合まで）で変化するのが、実行順を変えた後では必ず5となるため、最大50%の改善効果がある。

【0045】第四実施例として第二実施例の変形例を説明する。

【0046】図5のコマンド・スケジューラ32に連続領域でまとめたコマンド群を1つのコマンドとみなす手段を付加する方法がある。これは図9の(b)からも明らかのように、連続領域アクセス可能なコマンド群を1つのコマンドに再構成することで全体の処理コマンド数を減少し、まとめたコマンドについてはそのデータの大きさに関係なくディスク装置のプリフェッチ・バッファが有効利用できることで、プリフェッチ・ミスヒット（連続していないデータの場合、プリフェッチ・バッファにデータが無いことが多いためミスヒットが生じやすい）が避けられ、さらに、ディスク装置の回転待ち時間も回避したWrite処理の高速化が実現できる。

【0047】また、特にディスクアレイでは、Write時特有のパリティ更新に伴うRead処理を削減することができるため（データのアドレスが連続していない場合は、例えば、データが異なるとパリティディスクも異なる場合があるため、データごとにパリティディスクを読む必要が生じる）、Write処理の高速化が実現できる。

【0048】以上の説明より、連続領域アクセス可能性及び多重実行可能性双方を考慮したコマンドの並び替えも、図1に示す構成に第一実施例と第二実施例を用いることも可能である。また、第三実施例と第二実施例を用いることによっても第一実施例と第二実施例を用いる場合と同様な効果がもたらされる。また、第二実施例の代わりに第四実施例を用いることでさらに大きな効果がもたらされる。効果としては、実行順を変える前ではコマンド全てが各々Read/Writeアクセスする場合でも実行順を変えることで、連続領域アクセス可能なコマンドと多重実行可能なコマンドに整理されることからコマンド実行時間が短縮される。

【0049】

【発明の効果】本発明によれば、コマンドの実行時間の総和が短縮できる。

【0050】また、連続領域アクセス可能なコマンドをまとめて実行することで、ディスク装置のヘッドシークと回転待ちを短縮するという効果がある。

【0051】また、コマンドを多重実行することで、ディスク装置の稼働率を上げ、従来に対し最大50%のI/Oスループット改善効果がある。

【0052】さらに、キューのサイズを制限し、その中でコマンドの並び替えを行い、並び変えた順に実行するため、実行されないコマンドがいつまでもキュー内に残ってしまうコマンドの沈み込みを防止するという効果がある。

【図面の簡単な説明】

【図1】本発明に係るコマンド処理及び管理部のブロック図である。

【図2】本発明に係るフローチャートである。

【図3】コマンド・スケジューラの第一実施例を示すブロック図である。

【図4】本発明に係るフローチャートである。

【図5】コマンド・スケジューラの第二実施例を示すブロック図である。

【図6】本発明に係るフローチャートである。

【図7】コマンド・スケジューラの第三実施例を示すブロック図である。

【図8】本発明に係るフローチャートである。

【図9】各実施例によるコマンドの並び替え結果を示す説明図である。

【図10】ディスクアレイの基本構成を示すブロック図である。

【図11】ディスクアレイの基本構成を示すブロック図である。

【符号の説明】

10・・・上位計算機、11・・・アレイ・コントローラ、12・・・上位計算機I/F、13・・・タスク制御部、14・・・コマンド処理及び管理部、15・・・割込み処理部、16～20・・・ディスク装置I/F、21～25・・・ディスク装置、29・・・実行順最適化コマンド処理及び管理部、30・・・上位計算機コマンド受信部、31・・・アドレス変換部、32・・・コマンド・スケジューラ、33・・・コマンド待機キュー、34・・・コマンド並び替えキュー、35・・・キュー制御部、36・・・キューサイズ制限手段、37・・・実行順最適化制御部、38・・・連続領域アクセス可能性並び替え手段、39・・・多重実行可能性並び替え手段、40・・・キュー制御部、41・・・タイマー、42・・・ビット・パターン作成部、43・・・第一ビット・パターン読み取り部、44・・・第二ビット・パターン読み取り部、45・・・比較器、50・・・



13

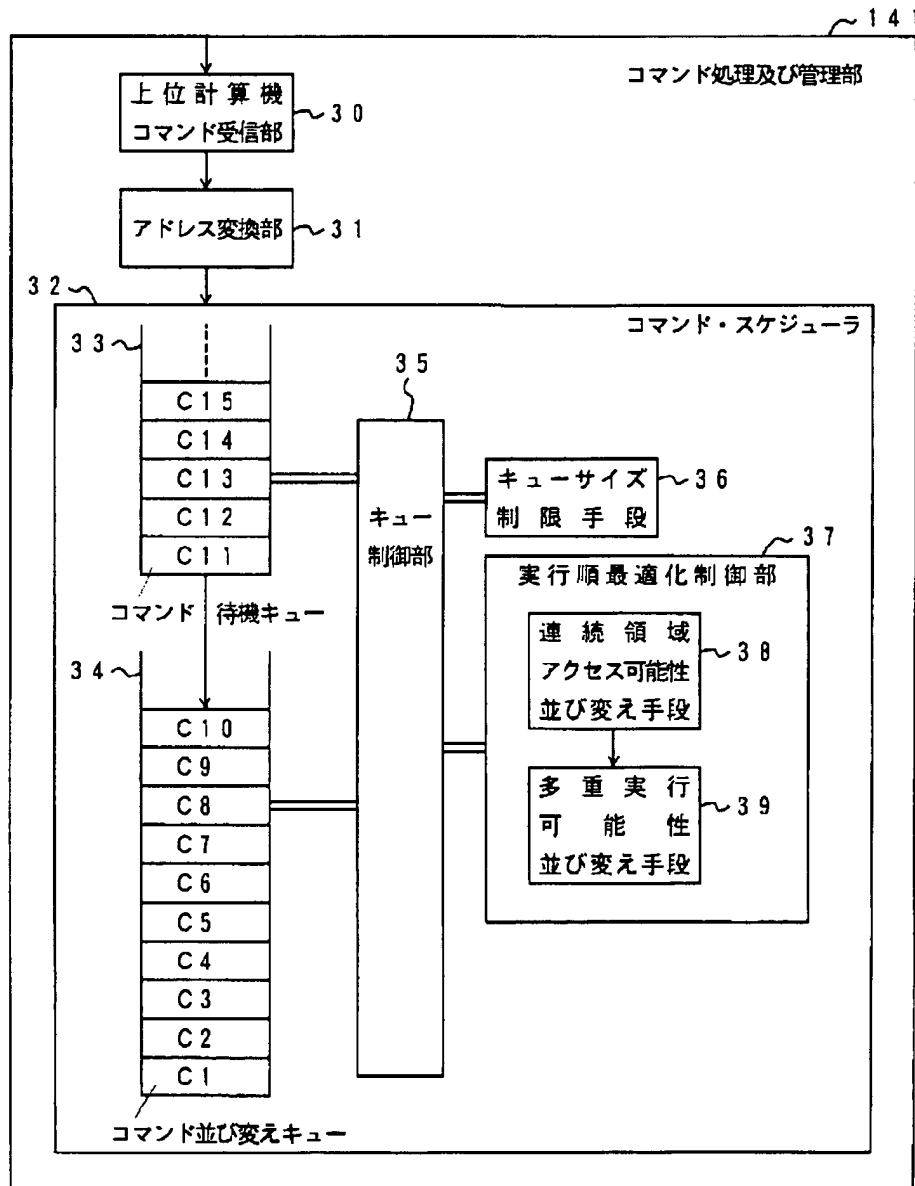
…キュー制御部、51… …連続領域判定アドレス作成手段、52… …第一アドレス読み取り部、53… …転送長読み取り部、54… …加算機、55… …第二アドレス読み取り部、56… …比較器、

14

60… …キュー制御部、61~70… …コマンド到着キュー、71~80… …フラグ、81… …コマンド実行キュー、82… …コマンド並び変えテーブル参照手段。

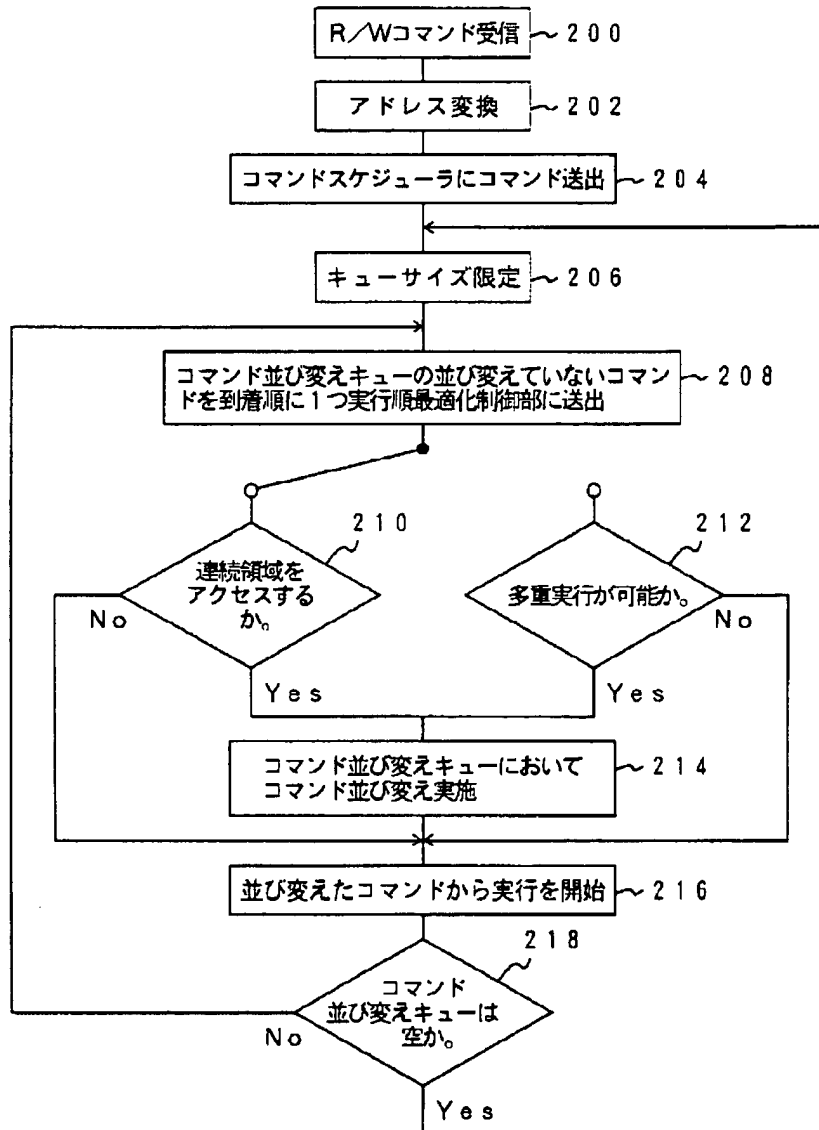
【図1】

図1 実行順最適化コマンド処理及び管理部のシステムブロック図



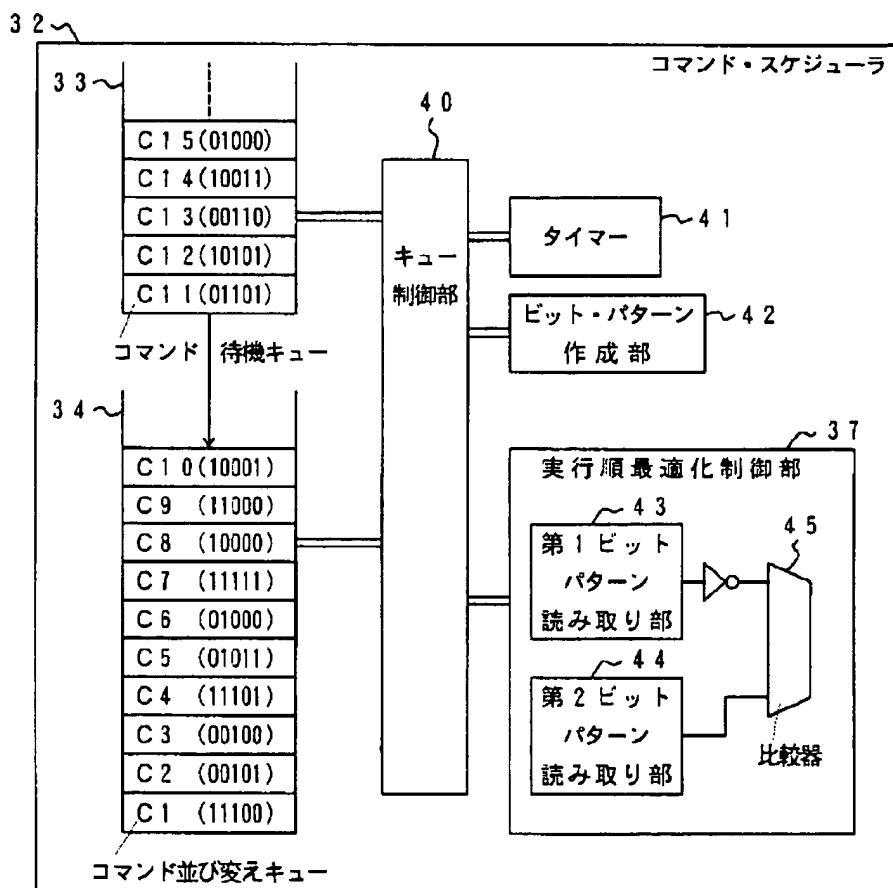
【図2】

図2 本発明の処理手段の基本動作を示すフローチャート



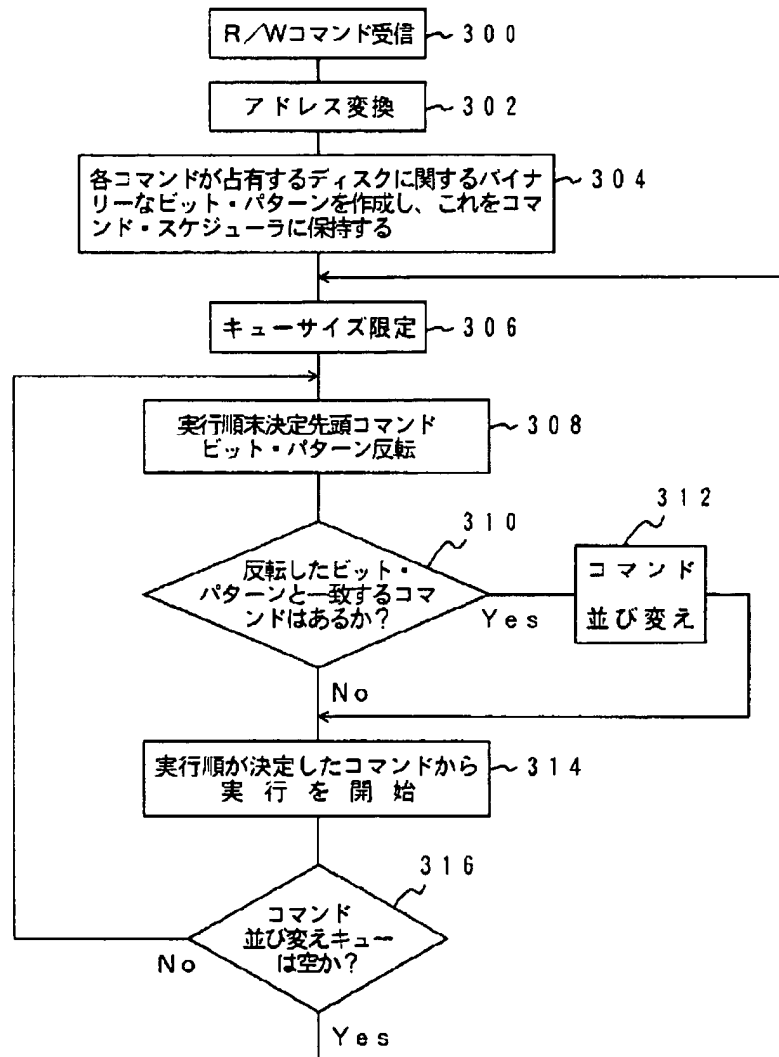
【図3】

図3 図1のコマンド・スケジューラの第一実施例を示すシステムブロック図



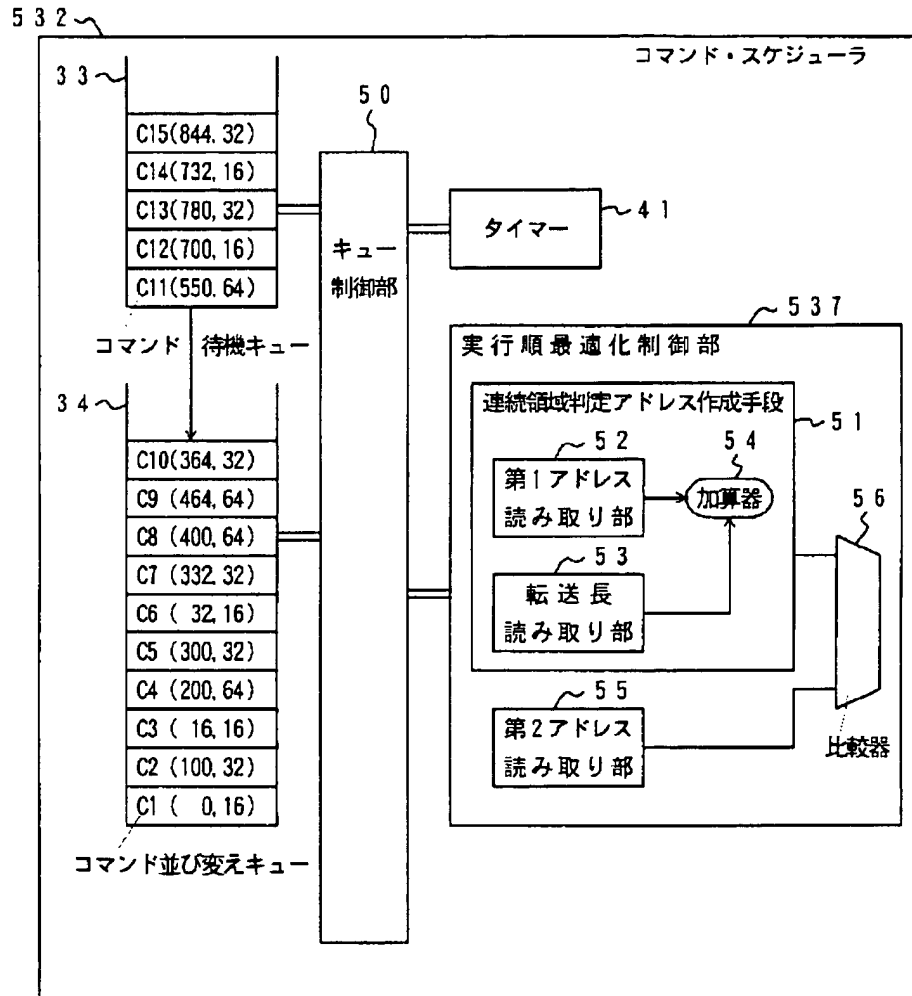
【図4】

図4 第一実施例の全体動作を示すフローチャート



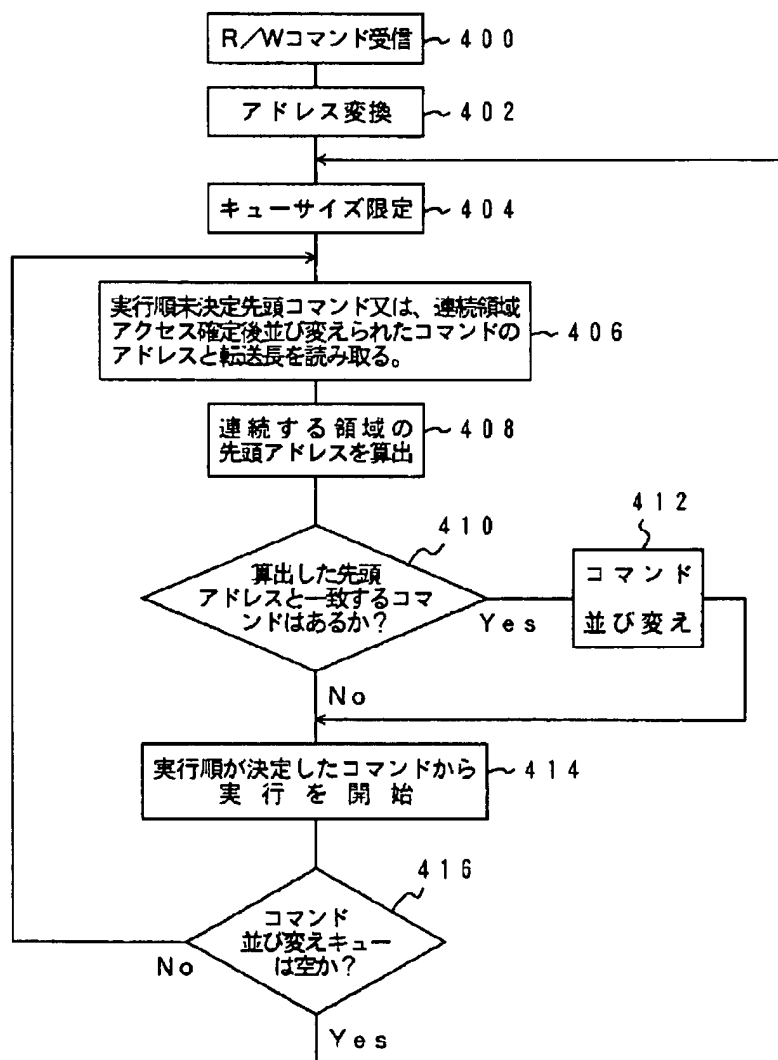
【図5】

図5 コマンド・スケジューラの第二実施例を示すシステムブロック図



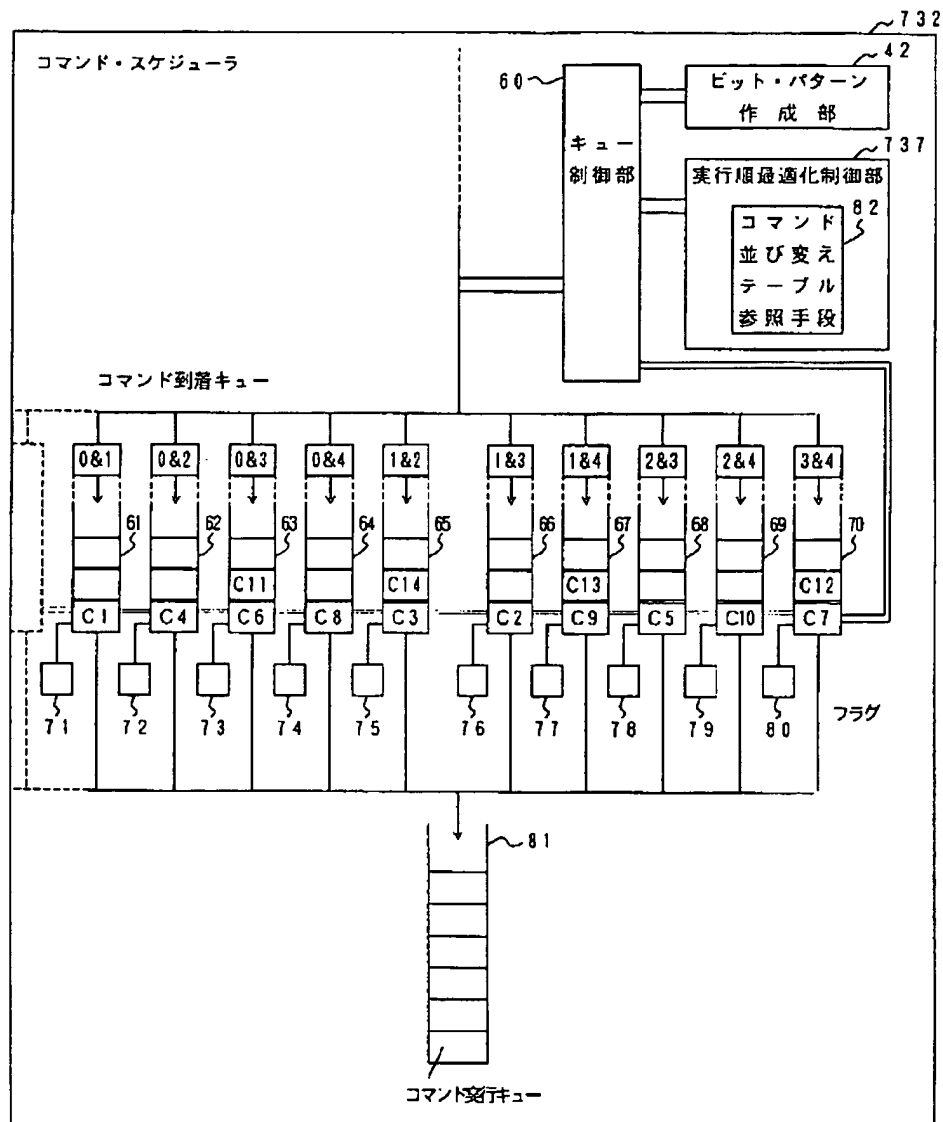
【図6】

図6 第二実施例の全体動作を示すフローチャート



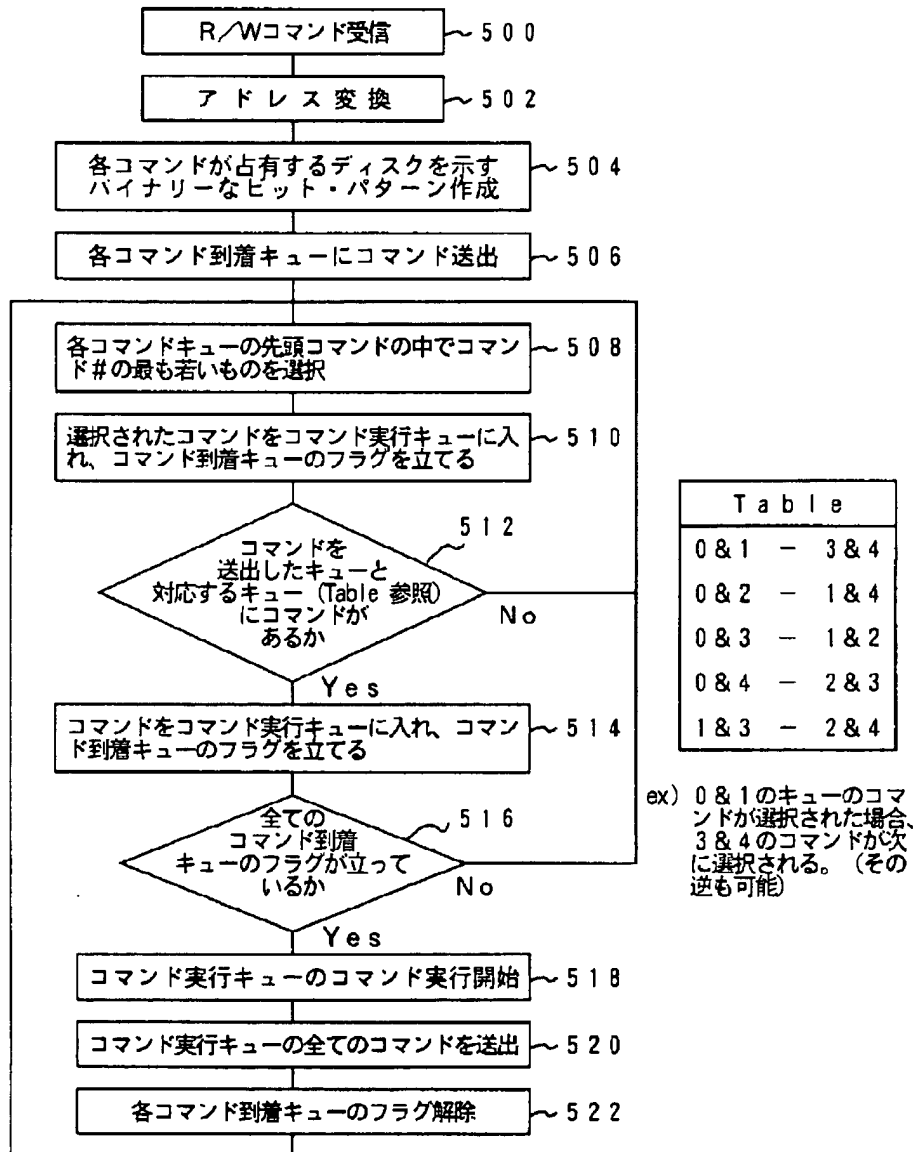
【図7】

図7 コマンド・スケジューラの第三実施例を示すシステムブロック図



【図8】

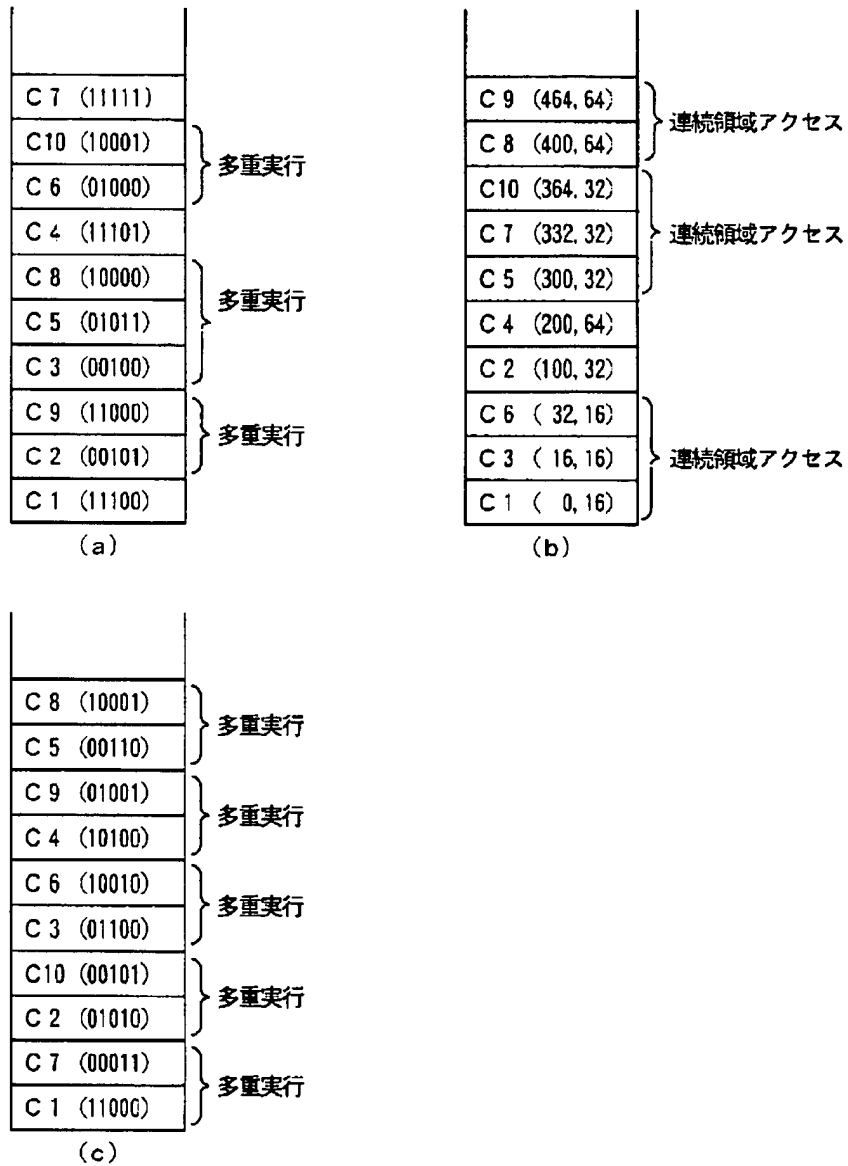
図8 第三実施例の全体動作を示すフローチャート





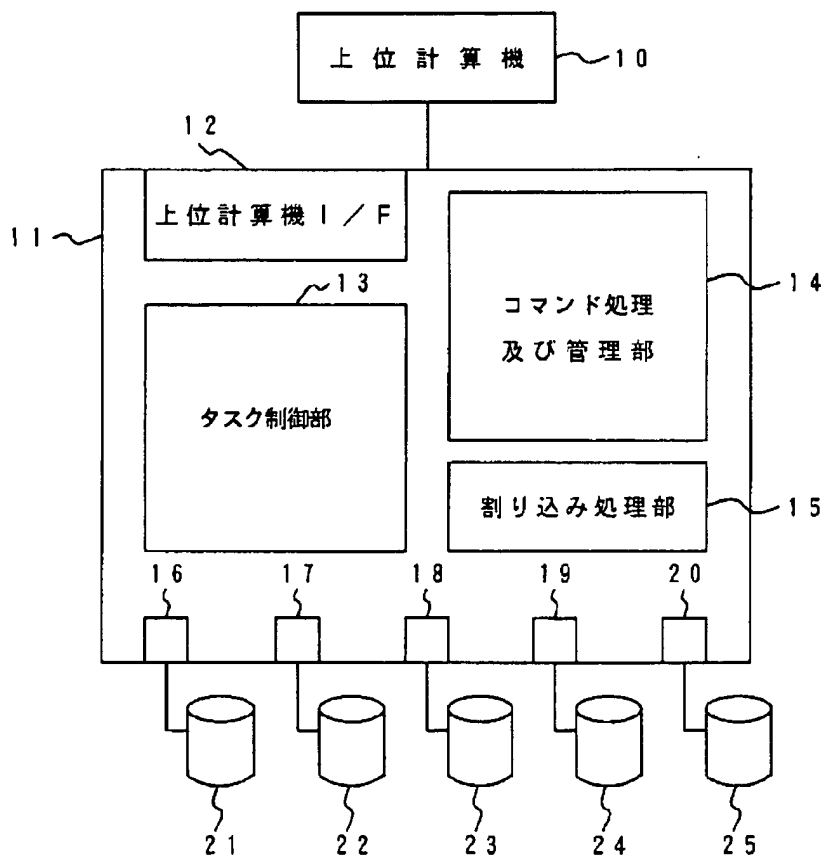
【図9】

図9 各実施例によるコマンドの並び変え結果を示す図



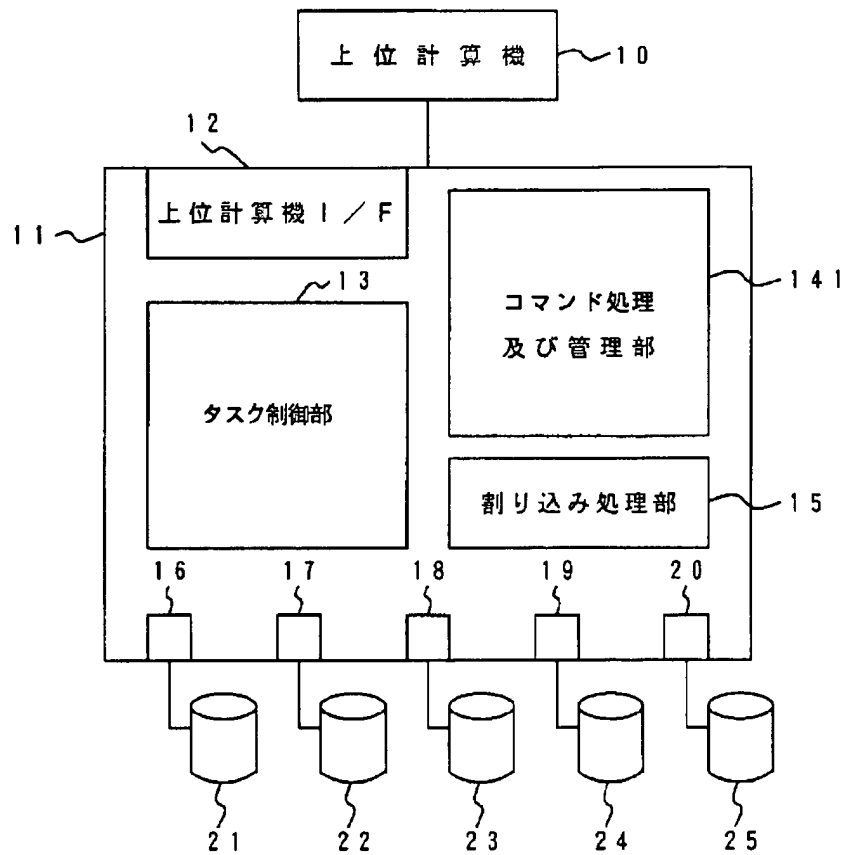
【図10】

図10



【図11】

図11



フロントページの続き

(72)発明者 大枝 高  
神奈川県横浜市戸塚区吉田町292番地 株  
式会社日立製作所マイクロエレクトロニク  
ス機器開発研究所内

(72)発明者 松並 直人  
神奈川県横浜市戸塚区吉田町292番地 株  
式会社日立製作所マイクロエレクトロニク  
ス機器開発研究所内